DOCUMENT RESUME

ED 060 062 TM 001 150

AUTHOR Wardrop, James L.
TITLE Towards a Broader Concept of Educational Assessment.
INSTITUTION Illinois Univ., Urbana. Center for Instructional Research and Curriculum Evaluation.
PUB DATE 14 Feb 72
NOTE 28p.; Paper presented at the Second Annual Mardi Gras Symposium, "Education: Instruction, assessment, and accountability," New Orleans, Louisiana, February 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Criterion Referenced Tests; *Educational Accountability; Educational Improvement; Educational Objectives; *Educational Testing; *Evaluation Methods; *Formative Evaluation; Measurement Goals; Measurement Techniques; Models; Performance Contracts; Performance Criteria; Standardized Tests; *Summative Evaluation

ABSTRACT

This paper illustrates how changing emphases and trends in education have led to a reformulation of notions of assessment and an awareness of the limitations of current measurement technology to deal with newer conceptualizations. Discussed are: 1) the antecedents of learning and the transactions which take place during the learning process, as well as with outcome measures; 2) the need to look at intentions; and 3) the concern with a plurality of goals and values as these relate to judgments based on assessment data. Problems highlighted include criterion-referenced measurement, the measurement of change, and the limitations of standardized tests as instruments for accountability. (Author/CK)

ED 060062

TOWARDS A BROADER CONCEPT OF EDUCATIONAL ASSESSMENT

James L. Wardrop
Center for Instructional Research and Curriculum Evaluation
University of Illinois at Urbana-Champaign

Prepared for

"Education: Instruction, Assessment, and Accountability"
Second Annual Mardi Gras Symposium
New Orleans, Louisiana

February 14, 1972

TM 001 150

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TOWARDS A BROADER CONCEPT OF EDUCATIONAL ASSESSMENT

James L. Wardrop
Center for Instructional Research and Curriculum Evaluation
University of Illinois at Urbana-Champaign

I find it very difficult to delineate a term like "educational assessment." At times, I am unable to differentiate this concept from "measurement." To assess in one sense is to measure. Tax assessors, after all, "assess" my property through what seems to me some mystical procedure of "measuring" its dollar value. A similar use is frequently found when educators speak of "assessing" pupil performance by employing some test or another. Here, too, assessment would seem to be synonymous with measurement. The National Assessment of Educational Progress, under the sponsorship of the Educational Commission of the States, seems to have been named on this basis.

At other times, one encounters "assessment" used as synonymous with "evaluation." In fact, one definition given in the third edition of Webster's New International Dictionary is "An appraisal or evaluation (as of merit)." Thus, we "assess" the effectiveness of a state's Title III program.

If I apply what seems to me to be some logical thinking, I come up with one statement to the effect that "assessment equals measurement" and another which indicates that "assessment equals evaluation." By my reasoning, I should now conclude that "measurement equals evaluation." But I know that ain't so! Besides, what do I do about something called "needs assessment," which according to a recent Educational Testing Service report (Educational Testing Service, 1971), is "universal....

Every state has conducted such a program, or is currently doing so, or is planning to recycle a completed one."?

Because of these various uses of the term "educational assessment," I feel that my first task in this presentation should be to attempt to delineate the concept of assessment in such a way as to differentiate it from measurement, on the one hand, and from evaluation, on the other. (For a somewhat different approach which deals with these same three terms, see Bloom's (1970) paper and subsequent discussions by Glass (1970), Guilford (1970), and especially Scriven's (1970) reaction.)

In an educational context, a rigorous, physical-science oriented conceptualization of measurement is generally inappropriate. Measurement basically involves the use of numerical values to represent attributes of objects. An attribute, in order to be measurable, must fit the specifications of a quantitative variable. Additionally, some unit of measurement must be established. Essentially, measurement may be defined as an "assessment of magnitude" (Jones, 1971).

We now have a definition of measurement as a particular kind of assessment of certain kinds of attributes. The implication is that we may "assess" in ways other than by measuring. There is something more involved in assessment than simply collecting and reporting measurements.

Perhaps two illustrations will help me to elaborate on the "something more." First, consider the National Assessment program mentioned earlier. On the surface, we think of National Assessment in terms

of the exercises administered and the results reported. If there were nothing more to National Assessment than developing, administering, scoring, and reporting results of exercises, it would better be called the "National Measurement" program. What more is there to the program which justifies the use of "assessment" in its name? Consider the activities which precede exercise development. Considerable effort is expended in developing statements of goals and objectives for each of the content areas in which measurement is to occur. Indeed, it is just this characteristic which is shared by the National Assessment project and the various statewide "needs assessment" programs. Prior to the specification, selection or development, and implementation of particular measurement techniques or strategies, considerable effort is devoted to making qualitative decisions about <u>what</u> to measure. Thus, assessment includes measurement, but additionally involves those qualitative and judgmental activities which go into determining what and how to measure. One <u>might</u> additionally include as a part of one's conception of assessment the processes of incorporating non-quantitative operations, of synthesizing the information obtained, and of making value-judgments about the attributes under investigation.

One essential component of all assessment activity is measurement. It is this feature which for me most sharply differentiates assessment from evaluation. I can point to some aspects of evaluation which involve neither measurement nor assessment. "All evaluation," as Stake (1969) once wrote, "deals explicitly with the worth of something." But it need not involve assessing that "something."

To summarize this attempt at clarification (which you may regard as obfuscation), we have three terms which differ among themselves in their specificity and their comprehensiveness. Measurement is the most specific, referring to our procedures for obtaining estimates of quantitative magnitude. Assessment includes, in addition to measurement, the processes through which goals and objectives are established, in which decisions are made about what to measure. Additionally, assessment allows the inclusion of qualitative information and the synthesis and interpretation of the information obtained. Evaluation, the most general of the three terms, sometimes includes assessment, but also allows for some approaches which simply do not fit with my use of the term "assessment."

With this incomplete, but hopefully sufficient, delineation of terms, I hope in the remainder of this presentation to show how our conceptualization of educational assessment has evolved during the past few years and to make some very speculative predictions about likely future trends.

The Eight-Year Study

A landmark in the development of modern concepts of educational assessment was the Eight-Year Study carried out by the Progressive Education Association during the 1930's. In particular, the work of the Evaluation Staff, under the leadership of Ralph W. Tyler, remains as one of the most important contributions ever made to educational evaluation. (I find the Tylerian view of "evaluation" corresponding quite closely

to my own notions of "assessment." In order to relate Tyler's work to the contemporary scene, I am going to use the former term, consistent with Tyler's own use, with the understanding that it is in the narrower sense of "assessment" that it should be interpreted.

In describing the purposes and procedures of the evaluation staff, Tyler (1942) presented the basic assumptions of the evaluation staff and described the general procedures they employed. Because these assumptions and procedures continue to have substantial influence on current evaluation practice, even more than theory, I am going to review them here.

There were eight particularly important assumptions made:

1. "Education is a process which seeks to change the behavior patterns of human beings."

2. "The kinds of changes in behavior patterns of human beings which the school seeks to bring about are its educational objectives."

3. "An educational program is appraised [assessed] by finding out how far the objectives of the program are actually being realized."

4. "Human behavior is ordinarily so complex that it cannot be adequately described or measured by a single term or a single dimension."

Because the next assumption is so very important, I am going to present Tyler's complete elaboration of it.

5. "The way in which the student organizes his behavior patterns is an important aspect to be appraised. There is always the danger that the identification of these various types of objectives will result in their treatment as isolated bits of behavior. Thus, the recognition that an educational program seeks to change the student's information, skills, ways of thinking, attitudes, and interests, may result in

> an evaluation program which appraises the development of each of these aspects of behavior separately, and makes no effort to relate them. . . . The way the student grows in his ability to relate his various reactions is an important aspect of his development and an important part of any evaluation of his educational achievement."

6. "The methods of evaluation are not limited to the giving of paper and pencil tests; any device which provides valid evidence regarding the progress of students toward educational objectives is appropriate."

7. "The nature of the appraisal influences teaching and learning."

8. "The responsibility for evaluating the school program belong[s] to the staff and clientele of the school."

(Tyler, 1942, pp. 11-14)

Given these assumptions (which could well have been written in 1972 rather than 1942), the general assessment procedure involved seven major steps: formulating objectives, classifying objectives, defining objectives in terms of behavior, suggesting situations in which the achievement of objectives will be shown, selecting and trying promising evaluation methods, developing and improving the more promising of these appraisal methods, and devising means for interpreting and using the results of the various instruments.

The efforts of Tyler and his staff continue to bear fruit. The decade of the 1960's saw an unprecedented exploitation of what has come to be called the "Tylerian model" of evaluation. (See, e.g., Glass [undated].) The specification of behavioral objectives and subsequent appraisal of an educational product in terms of the extent to which those objectives are in fact attained is perhaps the most pervasive of all evaluation strategies.

Popham (1969, p. 33) took perhaps the most optimistic view of the value of stating educational objectives in terms of learner behaviors when he wrote:

> We are at the brink of a new era regarding the explication of instructional goals, an era which promises to yield fantastic improvements in the quality of instruction.

It remained for Sullivan (1969) to spell out the implications of the specification of objectives for educational assessment. His treatment of the role of objectives in evaluation exemplifies what we might call the "neo-Tylerian" philosophy:

> Curriculum experts have emphasized the importance of precise instructional objectives for two primary purposes: planning instruction and assessing its effects.... Good instructional planning is based upon an assessment of the skills possessed by the intended student population, and the evaluation of instruction obviously must be based upon measurement of its outcomes.... The use of instructional objectives in evaluation can lead to educational improvement by resulting in the development and adoption of more effective curricula and by revealing the learning deficiencies of individual students and indicating appropriate treatments to overcome them.
>
> (Sullivan, 1969, pp. 80-81)

Other writers (e.g., Atkin, 1963; Eisner, 1967) have taken exception to the missionary zeal with which advocates of an objectives orientation to evaluation have presented their case. More recently, Scriven* has proposed what might be called a "radical alternative," which

*Scriven, M. "Goal-Free Evaluation," developed as a part of a planning project for the National Institute of Education and given only limited distribution.

he has christened "goal-free evaluation." In essence, he is proposing that externally-based assessment plans be developed independently of the stated goals and objectives of the project being evaluated. From the evaluator's point of view, orienting his assessment efforts around instructional objectives stated behaviorally is very seductive. His instrumentation task is made much easier if such specifications of desired learning outcomes are available. Under these conditions, it is a reasonably straightforward effort to develop appropriate measurement procedures to assess these outcomes. Such a strategy is of course best suited to the kind of evaluation in which a program is judged in terms of how well it achieves <u>its</u> goals, rather than, say, how well it achieves as <u>compared</u> to some other program or programs. (Scriven reminds us that we should not lose sight of the more important question: How <u>good</u> is the program? "Thus," he writes [1967, p. 53], "evaluation proper must include, as an equal partner of the measuring of performance against goals, procedures for the evaluation of the goals.")

Let us now look at some of the difficulties currently being encountered through the ardent pursuit of an objectives-based approach to educational assessment:

Recent pressures for accountability and the belief that there is a substantial group of pupils not being served by the educational establishment are among the influences which have made <u>performance contracting</u> a popular innovation. Basically, performance contracting is an arrangement in which an outside agency (the contractor) assumes the responsibility for some or all of a school or system's instructional

program in some or all content areas for some or all pupils. Thus, a school system might engage some company to provide basic reading instruction for all fifth-grade pupils whose level of reading achievement is some specified amount below their grade placement. An essential feature of performance contracts is the provision that the contractor's remuneration is to be based partially on gains in pupil performance on standardized tests. Typically, another portion of the remuneration is on the basis of pupil performance on the "criterion-referenced" measures incorporated in the contractor's instructional package. Here is the product-oriented assessment picture in clear and unmistakable focus. From the contractor's perspective, his livelihood depends directly, not on his ability to produce changes, but on his ability to produce measurable changes in pupil performance. Many measurement problems have been highlighted as a result of the performance contracting phenomenon (see, e.g., Feldmesser, 1971; Stake, 1971; Stake and Wardrop, 1971; and Wardrop, 1971 a and b), and a few alternative approaches have been suggested. If you choose to explore this topic, you will find few instances in which the appropriateness of a focus on outcome variables is questioned, many instances in which our earlier confidence in our ability to assess such outcomes is shaken, if not shattered.

From Product to Process

From the Eight-Year Study through the remainder of the pre-Sputnik era, the Tylerian approach dominated, essentially unchallenged because it worked so well. As seems to have been true of most aspects

of American education, the impact on assessment of the launching by the Soviets of the first earth satellite was unmistakable, albeit delayed and indirect. The late 1950's and early 1960's saw the beginnings of the large, national curriculum-development projects, projects whose initials are in all our vocabularies: PSSC, SMSG, BSCS, UICSM, etc. To varying degrees, each of these curriculum reform projects engaged in efforts to evaluate their products. Perhaps I can best capture the nature of the changing conceptualization of assessment during this period by quoting from Glass (undated, pp. 16-17):

> ...the men who had been involved in the "curriculum movement" of the late 1950's...carried with them into the late 1960's the baggage of objective achievement testing, taxonomies of objectives, the behavioral statement of instructional goals, etc.... [But] a model of evaluation was needed that would determine the value (worth, benefits) of activities as diverse as a mobile learning laboratory for children of migrant workers in Washington state, a computerized system of retrieving research information for teachers in Colorado, and a legitimate theatre for underprivileged children in New Orleans....

"It seems unlikely," Glass concludes, "that the Tylerian model of evaluation can grow to meet the new responsibilities of educational evaluation."

In turning to Glass for our denouement for the Tylerian approach, we have gotten a little ahead of ourselves in chronology. We were just beginning to consider the impact of the curriculum reform movement upon educational assessment.

It is generally true that the two major components of most assessment efforts in connection with the curriculum reform projects of the late 1950's and early 1960's were the Tylerian "objectives-oriented"

strategy and the tradition of experimental design borrowed from the researchers. The Tylerian model was especially seductive, because the specification of objectives, which is the sine qua non and greatest challenge of this approach, is a part of the curriculum development process itself, so that much of this part of the job was already done. The results of a number of the early assessment efforts along these lines were mixed. One of the most dramatic kinds of findings was that a curriculum package which appeared gratifyingly successful when employed under the careful supervision of the development staff and under carefully controlled conditions of administration would appear to be "no better" if not in fact worse than existing offerings when subjected to field-test conditions. Why this apparent anomaly should have occurred seems obvious to us now, a decade later. At the time, however, it took considerable exploration to uncover the fact that often what teachers were doing in the classroom was essentially independent of the materials they used. Even though a teacher were given SMSG mathematics materials to use, he would continue in his classroom behavior to act as if he were using the traditional mathematics materials. I do not, in this presentation, want to get into a consideration of some of the corrective strategies devised to deal with the problem. In the context of educational assessment, the point is that some people began to realize that assessments of educational programs must attend not only to pupil performance outcomes but also to what happens during the instructional sequence itself; what Stake (1967) has called "transactions" and Stufflebeam (1969; Stufflebeam et al., 1971) refers to as "process."

Do not misunderstand me. I do not mean to imply that no one had ever before considered such transactions as a part of educational assessment; nor do I wish to indicate that assessments of process variables have ever predominated (or ever should, for that matter). Rather, I want to indicate that the relative emphasis on so-called process variables increased markedly at that time.

By way of illustration, consider the study by Anderson (1968), which reported on an evaluation which employed the comparative field experiment methodology. In describing the conceptual background of the study, he dealt with some of the problems which indicate the need for assessing transactional events:

> There are no procedural features of lessons that are invariably associated with greater student achievement. Neither small steps, nor active responding, nor immediate feedback, nor a warm classroom climate, nor a sequence from concrete to abstract, nor the provision for self-direction and self-pacing, nor multi-media stimulus bob-bardment--singly or in the aggregate--guarantee successful instruction.
>
> (Anderson, 1968, pp. 3-4)

In his study, Anderson did in fact collect considerable data on the manner in which the treatment (a self-instructional program on population genetics) was implemented. Among his data-collection procedures were teacher logs, teacher questionnaires, and pupil questionnaires.

Among the analyses of pupil achievement was one which explored "achievement as a function of the teacher." Reporting on these analyses, Anderson (1968, p. 17) noted that there was "enormous variation in the ways teachers used the program. Some teachers did not allow any class

time for students to study the program while, at the opposite extreme, there were teachers who used the program, and only the program, to teach population genetics." When teachers were classified according to how they assigned the program, some interesting differences in pupil achievement were noted. (The details of Anderson's findings are not of concern here, so I will leave it to you to read the original report if you are interested.)

In his study, Anderson chose to assess transactions through various self-report techniques. One of the alternatives for assessing classroom transactions is the use of observers who complete some classroom observation form. Considering the place of classroom interaction information in assessment, Stake (1970a, p. 2) noted that "Even people who expect that the particular ways a teacher and child interact have little effect on what he learns are likely to want to keep track of classroom conditions within which 'more crucial' forces acted.... Most [evaluation report readers] look for some data on the ways in which teacher and students interacted."

A few paragraphs later, Stake (ibid., p. 5) concluded:

> The disgraceful aspect of the evaluation of thousands of educational innovations in the last decade is not that we do not know what the children learned, but that we do not know how and what the teachers taught. The saying goes, "What the child has not learned, the teacher has not taught." But much of what has been learned cannot be known, but how the learning opportunity has been arranged can be. And that information can be of high priority. Neither an understanding of what the curriculum has been or what should be tried next time is possible without data on the teaching methods. In some evaluation studies the most valuable data will be those gathered by a classroom observation system.

In spite of some extensive work on systems for observing and classifying classroom interactions--which has, by the way, resulted in the development of some 70 or so different observational systems--it is still true, as Rosenshine and Furst (in press) noted, that:

> Just as it is relatively easy to develop new observational systems, it has been fairly easy for educators to develop lists of teaching skills. Unfortunately, the teaching skills, just like the observational systems, are seldom validated against measures of student growth.

A somewhat different perspective on transactions has begun to emerge very recently with the increasing popularity of what has been called "open education." In such a setting, where individuals and small groups of pupils pursue unique learning tracks with but minimal prescription, how does one even begin to assess the effectiveness of the overall approach? Wolf (1971) suggests that it is the nature of the transactions, encounters, and the process of learning which provide the components which ultimately differentiate open education from the more traditional "teacher-centered" orientations. This conceptualization, supported by Stake's (1967) and Scriven's (1967) arguments for the importance of transactions, Eisner's (1969) treatment of "expressive objectives," and Arnstine's (1964, 1967) argument for transactions that have "aesthetic quality," leads him to conclude that "transactions are part of the learning process and therefore possess an intrinsic value by themselves." (Wolf, 1971, p. 39). Unfortunately, he argues, the currently relied-upon indicators of such transactions are inadequate to the task.

In spite of Anderson's exemplary study, in spite of a plethora of classroom observation systems, in spite of Wolf's focus on transactions, in spite of what seems to have amounted to virtually a quantum-jump in emphasis on assessing process variables, our methodology has for the most part lagged behind. How can we capture the essential quality of a classroom event? How can we describe, assess, summarize, synthesize, and report in any meaningful way just what has transpired in any one event and what its implications are for the total educational process? Lacking answers to these questions, I am going to proceed now to what I see as the more recent change in our conceptualization of assessment.

Assessing the Context of Education

In their introduction to State Educational Assessment Programs, Dyer and Rosenthal (1971, p. ix) note three impacts on educational assessment during the middle 1960's:

> The first was the formation in 1964 of the Exploratory Committee on the Assessment of Progress in Education, which eventuated in the National Assessment program now underway. The second event was the enactment of the Elementary and Secondary Education Act of 1965, which included a requirement that school systems assess by objective means the effects on student achievement produced by federally funded programs for the educationally deprived. The third was the publication in 1966 of the Coleman report on Equality of Educational Opportunity, which attempted to assess, again in terms of measured pupil achievement, the quality of service the schools were supplying to various segments of the population.

The immediate impact of these developments on the field of educational assessment was a renewed concern for measuring pupil performance. The long-range impact seems to have been to emphasize the

need for a broader kind of assessment. In particular, evaluation reports from the earlier projects funded under ESEA were essentially useless for assessing the total impact of this act on American education. The data collected were simply not amenable to aggregation and interpretation in a way which would be useful to administrators or legislators at the national level. Partially as a result of experiences of this sort, officials in the U.S. Office of Education have tried several different strategies for evaluating programs of national scope. One approach has been to contract with organizations or individuals specifically to evaluate these national programs apart from the individual projects. For many reasons, these efforts have most commonly generated an inordinate amount of mutual antagonism for the parties involved. The other most visible attempt to deal with the problem is (or was) the Federal-State Joint Task Force on Evaluation, better known as the "Belmont Project." This project, noble in its conception, chaotic in its implementation, now seems likely to become an outstanding example of a project gone awry.

We have seen earlier in this presentation that Tyler's rationale for focusing assessment efforts on changes in pupil performance continues, and rightly so, to be a pervasively adopted one. We have also seen the emergence of a concern for how a program is implemented, for what goes on in the classroom, for what I have called "process variables." Now I want to try to make a case for the appearance of a quite different kind of emphasis in educational assessment. The time perspective is too short for me to speak with much confidence about this trend--if indeed it be a trend--but I am going to try anyway.

My contention is that the new demands on educational assessment as a result of the growth of such broad educational intervention programs as Project Headstart, Title I of the 1965 ESEA, and Project Follow-Through could not be met by applying the models and strategies based on earlier conceptualizations. David Cohen (1970, p. 213) has described some of the differences between these newer programs and the traditional objects of educational evaluations:

> (1) they are social action programs, and as such are not focused narrowly on teachers' in-service training or on a science curriculum, but aim broadly at improving education for the disadvantaged; (2) the new programs are directed not at a school or a school district, but at millions of children, in thousands of schools in hundreds of school jurisdictions in all the states; (3) they are not conceived and executed by a teacher, principal, a superintendent, or a researcher--they were created by the Congress and are administered by federal agencies far from the school districts which actually design and conduct the individual projects.

Perhaps because such programs as Cohen has described involve the allocation of a substantial portion of a finite pool of resources, some writers have argued that one important role of evaluators lies in questioning the legitimacy and value of the objectives of the program being evaluated. Although it takes us beyond the limits of "assessment" (but not of "evaluation"), one aspect of a recent evaluation (Stake and Gjerde, 1971) exemplifies this newer approach. The evaluation of the Twin City Institute for Talented Youth (TCITY) dealt more explicitly with project goals than is usually the case. In the words of its director, "The primary objective of the Twin City Institute is to create an educational program that has strong academic and social appeal for

students who possess a variety of artistic, language, scientific and leadership talents." (Stake and Gjerde, 1971, p. 4). He goes on to talk about "an atmosphere of freedom," an emphasis on inquiry, openness, creativity, and the "humanizing" aspects of learning.

One unique component of this evaluation report is the inclusion of "An Adversary's Statement" (Denny, 1971, p. 27). Among his criticisms is the following:

> How costly is this Institute? Dollar costs are sufficient to give each group of six students $1,000 to design and conduct their own summer experience. Over 100 Upward Bound students could be readied for their college careers...About twenty-five expert curriculum specialists could be supported for half a year to design and develop new curricula for the high school.

Now, I prefer not to call <u>this</u> aspect of evaluation "assessment." (Remember, earlier I indicated that evaluation is something more than assessment. You should have been picking up some cues as to the nature of the differences as we go along.) Yet the approach reflected in Denny's statement can influence the nature of what is done in the name of assessment. Specifically, what seems to be happening with somewhat greater frequency now than in the past is that evaluators are addressing themselves to the issue of goals and values, especially in the context of competition for resources. Recent studies by Gooler (1971) and McQuarrie (1971) represent explorations of alternative methodologies for quantifying judgments of value and priority. The relationship of this concern with values and priorities to the broadening of our concept of educational assessment, not the particular methodological approaches that might be utilized, is of concern here.

Essentially, the argument begins with an assertion that an important component of educational assessment should be a consideration of the intents of the object being evaluated as those intents are related to the value (priority) structures of important reference groups. Consider such reference groups as, say, parents, school board members, various community organizations, school administrators, teachers, and pupils. How highly do these groups value what a particular educational program seeks to accomplish? In particular, what other legitimate educational goals are they willing to sacrifice in order to support this program? The questions of cost raised by Denny in the TCITY evaluation (Stake and Gjerde, 1971) are in fact amenable to this kind of assessment. It is certainly within the realm of possibility to undertake an assessment of relative priorities of reference groups with respect to the alternatives suggested by Denny (and others he did not consider). Some creativity would be needed in developing appropriate assessment strategies, but we can certainly get some information about how the various groups would choose among the TCITY approach, with its focus on talented youth; a project to prepare Upward Bound students for college; the support of expert curriculum specialists to design and develop new high school curricula; or some other educational program. Stake (1970b) --that name does keep coming up--has made a plea for incorporating such data into our conceptualization of educational assessment. His colleagues and students, at least, are attempting to honor that plea.

Values and priorities are but one aspect of the context of education. Other aspects of the context in which formal educational

programs occur have likewise been receiving increased emphasis recently. The development of the "CIPP" model for evaluation (Stufflebeam, 1969) --with "C" for context, "I" for input, "P" for process, and "P" for product--and the presentation of the so-called "countenance model" (Stake, 1967) have focused our attention on the importance of considering contexts when assessing educational activities. Of course, it should be remembered that the "accreditation model" of evaluation represented by such groups as the North Central Association has for well over half a century focused almost exclusively on "context" variables, but with a much more limited methodology than is being advocated here. (For a more complete description of the accreditation approach, see Stake [1970b] or especially Glass [undated].)

Perhaps I ought to elaborate on what I mean by "context variables." My notion is a broad one with many levels of meaning. It includes, but seems not to coincide precisely with Stake's (1967) "antecedents" or Stufflebeam's (1969) "context." At the most general level, context variables refer to the social, philosophical, historical, anthropological, economic, and political milieus in which educational programs function. Yes! All of these--and I may have left out a few--are a part of the context of education. Only a subset of them is at all amenable to "assessment," and a still smaller subset is included in the domain of feasibility. The kind of value or priority assessment discussed earlier is <u>one</u> approach to assessing <u>one</u> manifestation of this context--the priorities of selected reference groups. Another approach would emphasize the relationship of a program's transactions to the societal

expectations about, say, "cooperation versus competition." That is, one could assess attitudes of appropriate reference groups concerning the extent to which cooperation among individuals should (a value statement) be fostered in a particular educational program, then observe the extent to which the classroom behaviors of pupils match or deviate from these expectations.

As another example, one might consider the combination of, say, the political-social-economic contexts in contemporary America as the basis for assessing the federally-based social action programs in education. Cohen's (1970) discussion presents a superb rationale for such an approach.

At another level, the physical setting and facilities in which a program is carried out are a part of the "context." Assessments by regional or national accrediting agencies place considerable emphasis on these attributes: average class size; number and types of books, periodicals, etc. in the library; and currency of available textbooks to mention just a few examples.

Another category of "context variables" seems to fit rather more closely what Stake (1967) called "antecedents." In this category, one finds such attributes as level of training of the teachers; achievement history of the pupils, as well as their aptitudes, attitudes, and motives; and other enabling (or disabling) characteristics which might play an essential role in determining the success (or failure) of a program. Some of these variables have long been a part of the evaluations conducted by accrediting agencies (see, in addition to the

references cited earlier, Davis [1945]). Most evaluation activities, however, have tended to underplay their importance. One consequence of recent writings by such evaluation theorists as Stake (1967) and members of the Phi Delta Kappa National Study Commission on Evaluation (Stufflebeam, et al., 1971) has been at least to remind us that such variables often bear an important relationship to the perceived success or failure of a program. This increased attention to context represents, to me, the major current thrust in educational assessment.

Perspective and Prospective

In the preceding sections, I have traced the development of educational assessment through three stages. In the beginning was Ralph Tyler and the commandments of product-oriented assessment.

Many years later came the deluge of national curriculum projects, followed by the process-servers, with their faith that we could understand outcome variables if only we were to look at "what went on." The most recent article of faith takes context as its text. We will finally understand outcome variables if only we consider the context in which the processes occur. I could put everything into one multiple-choice question. (Isn't that where it's really at, after all?)

The question:

Which of the following best describes the important perspectives to be considered in educational assessment?

a. Outcome measures based on instructional objectives stated in behavioral terms.

b. Process measures describing the ways in which instructional programs are in fact implemented.

c. Context measures addressing themselves to general issues of value and priority and particular issues of environmental setting.

d. None of the above.

e. All of the above.

The poet Wallace Stevens once wrote a poem entitled, "Thirteen Ways of Looking at a Blackbird." While I do not mean to suggest any commonality between educational assessment and blackbirds (although various other metaphors relating educational measurers, assessors, or evaluators to certain kinds of birds come readily to mind: the owl, the falcon, or possibly the albatross, depending on one's orientation), Stevens' poem does provide a useful analogy for me. In this presentation, I have illustrated one way of looking at educational assessment. The perspective I chose has resulted in my selecting certain attributes to observe and describe. More importantly, perhaps, there are other attributes which one should observe and describe, given some other perspective. Educational assessment is complex and multi-faceted. No assessment of assessment can capture all its dimensions. Think about that. Then re-state it like this: any educational activity is complex and multi-faceted. No assessment of a program can capture all its dimensions. Think about that.

Then think about this: any educational assessment represents a compromise. We assess to find out "the way things are." Then, and only then, can we rationally decide if things are as they should be. But--every assessment is incomplete. It reflects many decisions made

along the way. It represents but one perspective on "the way things are." One perspective is not enough. Many perspectives are needed.

In concluding this presentation, I considered using the well-known story of the blind men trying to describe an elephant, but that struck too close to home. I also considered the story related by Messick (1970) about the rabbinical student named Ezekiel, but his use of it was much more appropriate than mine would be. I chose instead to turn to another poem by Wallace Stevens, "Connoisseur of Chaos," which begins:

> A. A violent order is disorder; and
> B. A great disorder is an order. These
> Two things are one. (Pages of illustrations.)

What I have presented here is thus aptly described. Need I say more?

REFERENCES

Anderson, R. C. "The Comparative Field Experiment: An Illustration from High School Biology." In Proceedings of the 1968 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1968, 3-30.

Arnstine, D. "Shaping the Emotions: The Source of Standards for Aesthetic Education." The School Review, 1964, 72, 239-271.

Arnstine, D. Philosophy of Education. New York: Harper and Row, 1967.

Atkin, J. M. "Some Evaluation Problems in a Course Content Improvement Project," Journal of Research in Science Teaching, 1963, 1, 129-132.

Bloom, B.S. "Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment." In M. Wittrock and D. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970, 25-50.

Cohen, D. K. "Politics and Research: Evaluation of Social Action Programs in Education," Review of Educational Research, 1970, 40, 213-238.

Davis, C. O. A History of the North Central Association. Ann Arbor, Michigan: North Central Association of Colleges and Secondary Schools, 1945.

Denny, T. "An Adversary's Statement." In R. Stake and C. Gjerde, An Evaluation of TCITY. Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign, 1971, pp. 27, 14. (Mimeo.)

Dyer, H. S., and Rosenthal, Elsa. "An Overview of the Survey Findings." In State Educational Assessment Programs. Princeton, N.J.: Educational Testing Service, 1971, pp. ix-xix.

Educational Testing Service. State Educational Assessment Programs. Princeton, N.J.: Educational Testing Service, 1971.

Eisner, E.W. "Educational Objectives: Help or Hindrance?" School Review, 1967, 75, 250-266.

Eisner, E. "Instructional and Expressive Objectives: Their Formulation and Use in Curriculum. In Instructional Objectives. (AERA Monograph Series on Curriculum Evaluation, No. 3). Chicago: Rand-McNally, 1969, 1-18.

Feldmesser, R. A. "Measurement problems in performance contracting." NCME Measurement News, 1971, 14 (3), 6-7.

Glass, G. V. "Comments on Professor Bloom's Paper." In M. Wittrock and D. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970, 56-61.

Glass, Gene V. "The Growth of Evaluation Methodology." Boulder, Colorado: Laboratory of Educational Research, University of Colorado, undated. (Mimeo.)

Gooler, D. D. Strategies for Obtaining Clarification of Priorities in Education. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1971.

Guilford, J. P. "Comments on Professor Bloom's Paper." In M. Wittrock and D. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970, 62-66.

Jones, Lyle V. "The Nature of Measurement." In R. Thorndike (ed.), Educational Measurement (Second Edition), Washington, D. C.: American Council on Education, 1971, 335-355.

McQuarrie, D. M. The Perceived Dimensions of Value Underlying Broad Educational Objectives. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1971.

Messick, S. Comments recorded in M. Wittrock and D. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart, and Winston, 1970, 177-178.

Popham, W. J. "Objectives and Instruction." In Instructional Objectives. (AERA Monograph Series on Curriculum Evaluation, No. 3). Chicago: Rand-McNally, 1969, 32-52.

Rosenshine, B., and Furst, N. "The Study of Teaching in Natural Settings Using Direct Observation." In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand-McNally, in press.

Scriven, M. "The Methodology of Evaluation." In Perspectives of Curriculum Evaluation. (AERA Monograph Series on Curriculum Evaluation, No. 1). Chicago: Rand-McNally, 1967, 39-83.

Scriven, M. "Science and Evaluation: Reflections on Professor Bloom's Paper." In M. Wittrock and D. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970, 50-56.

Stake, R. E. "The Countenance of Educational Evaluation." Teacher's College Record, 1967, 68, 523-540.

Stake, R. E. "Generalizability of Program Evaluation: The Need for Limits." EPIE Forum, 1969, 2 (5), 39-41.

Stake, R. E. "The Decision: Does Classroom Observation Belong in an Evaluation Plan?" In Classroom Observation. (AERA Monograph Series on Curriculum Evaluation, No. 6). Chicago: Rand-McNally, 1970(a), 1-5.

Stake, R. E. "Objectives, Priorities, and Other Judgment Data." Review of Educational Research, 1970(b), 40, 181-212.

Stake, R. E. "Testing hazards in performance contracting." Phi Delta Kappan, 1971, 52, 583-589.

Stake, R. E., and Gjerde, C. "An Evaluation of TCITY, The Twin City Institute for Talented Youth." Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign, 1971. (Mimeo.)

Stake, R. E., and Wardrop, J. L. "Gain score errors in performance contracting." Reading Research Quarterly, 1971, 6, 323-325. Also in Research in the Teaching of English, 1971, 5, 226-229.

Stevens, Wallace. Poems. New York: Vintage Books, 1959.

Stufflebeam, D. L. "Evaluation as Enlightenment for Decision Making." In Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Washington, D. C.: Association for Supervision and Curriculum Development, National Education Association, 1969, 41-73.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., and Provus, M. M. Educational evaluation and decision making. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1971.

Sullivan, H. J. "Objectives, Evaluation, and Improved Learner Achievement." In Instructional Objectives. (AERA Monograph Series on Curriculum Evaluation, No. 3). Chicago: Rand-McNally, 1969, 65-90.

Tyler, R. W. "Purposes and procedures of the evaluation staff." In E. R. Smith and R. W. Tyler (eds.), Appraising and Recording Student Progress. New York: Harper & Brothers, 1942, 3-34.

Wardrop, J. L. "Providence to New Century to Chance: Double Play or Double Cross?" Educational Product Report, 1971a, 5 (2), (Whole number 38), 1-5.

Wardrop, J. L. "Measuring Change: Some Problems and Solutions." Paper presented at the National Reading Conference, December 1971b. (Mimeo.)

Wolf, R. L. "Making Education Accountable to the Learner: A Framework for Evaluation." Journal of Research and Development in Education, 1971, 5, 37-48.